

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:05:58 ; Search time 1571 Seconds
(without alignments)
324.885 Million cell updates/sec

Title: US-10-085-108-21_COPY_711_731

Perfect score: 21

Sequence: 1 AAGCTGATGATCACCAGGG 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	87.6	686	29	CC270269 CH261-1N2
2	18	85.7	809	29	B2721194 PUCFH73TD
C 3	17.8	84.8	307	9	AA894450 oe79f07.s
C 4	17.8	84.8	339	9	AI572687 tr78b10.x

C 5	17.8	84.8	380	28	BH323391
C 6	17.8	84.8	545	14	CA941354
C 7	17.8	84.8	592	9	AI816423
C 8	17.8	84.8	671	29	AG121419
C 9	17.8	84.8	708	29	AG186976
C 10	17.8	84.8	913	13	BH245702
C 11	17.8	84.8	963	10	BG119393
C 12	17.8	84.8	1201	9	AL523235
C 13	17.4	82.9	565	28	AZ072191
C 14	17.4	82.9	914	10	BE729033
C 15	17	81.0	543	9	AI818091
C 16	17	81.0	733	14	CB983939
C 17	17	81.0	763	14	CD101012
C 18	16.8	80.0	321	9	AA967543
C 19	16.8	80.0	341	9	AA118769
C 20	16.8	80.0	348	29	AG236818
C 21	16.8	80.0	359	10	BE861602
C 22	16.8	80.0	364	28	BH056156
C 23	16.8	80.0	369	9	AW229797
C 24	16.8	80.0	409	9	AA636853
C 25	16.8	80.0	420	10	BE181117
C 26	16.8	80.0	457	9	AI225508
C 27	16.8	80.0	457	10	BF322594
C 28	16.8	80.0	497	10	BG276054
C 29	16.8	80.0	509	9	AW230744
C 30	16.8	80.0	512	13	BU583368
C 31	16.8	80.0	522	10	BG078216
C 32	16.8	80.0	523	13	BQ356299
C 33	16.8	80.0	526	13	BQ287718
C 34	16.8	80.0	543	9	AA798972
C 35	16.8	80.0	544	4	BX517147
C 36	16.8	80.0	554	9	AA543306
C 37	16.8	80.0	556	9	AA543338
C 38	16.8	80.0	563	14	W19670
C 39	16.8	80.0	581	14	CB271143
C 40	16.8	80.0	583	28	AQ047644
C 41	16.8	80.0	587	13	BQ444442
C 42	16.8	80.0	593	13	BQ265035
C 43	16.8	80.0	600	12	BI987638
C 44	16.8	80.0	604	9	AA544735
C 45	16.8	80.0	640	29	BZ421605

ALIGNMENTS

RESULT 1	CC270269	CH261-1N21 Sp6.1	686 bp	DNA	linear	GSS 13-MAY-2003
LOCUS	CH261-1N21	Sp6.1	686 bp	DNA	linear	GSS 13-MAY-2003
DEFINITION	CH261-1N21	Sp6.1	686 bp	DNA	linear	GSS 13-MAY-2003
ACCESSION	CC270269	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
VERSION	CC270269.1	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
KEYWORDS	GSS.	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
SOURCE	Gallus gallus (chicken)	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
ORGANISM	Gallus gallus	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
REFERENCE	1 (Bases 1 to 686)	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
AUTHORS	Kremizki, C., Higginsbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
TITLE	Gallus gallus BAC End Reads	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
JOURNAL	Unpublished	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
COMMENT	Contact: Richard K. Wilson	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	Genome Sequencing Center	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	Washington University School of Medicine	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	Email: submissions@wustl.edu	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	Insert Length: 182000 Std Error: 0.00	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	Seq primer: Sp6 ATTAGGTGACATATAG	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	Class: BAC ends	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	High quality sequence start: 68	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003
	High quality sequence stop: 498.	CH261-1N21	686 bp	DNA	linear	GSS 13-MAY-2003

```

FEATURES
  source
    Location/Qualifiers
      1. .686
        /organism="Gallus gallus"
        /mol_type="genomic DNA"
        /strain="Red Jungle Fowl"
        /db_xref="taxon:3031"
        /clone="CH261-1N21"
        /sex="Female"
        /cell_line="CH261"
        /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
        CH261 Female Chicken library - For library and clone
        ordering information: http://www.chori.org/bacpac"
      196 a 147 c 117 g 226 t
      BASE COUNT
      ORIGIN
        Query Match 87.6%; Score 18.4; DB 29; Length 686;
        Best Local Similarity 95.0%; Pred. No. 3.4e+02;
        Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

      Qy 2 AGCTGATTGATGACCCAGGG 21
      Db 524 AGCTGATTGAAGACCCAGGG 543

RESULT 2.
BZ721194
LOCUS
DEFINITION
  BZ721194 809 bp DNA linear GSS 24-FEB-2003
  PUCFH73TD ZM 0.6 1.0 kb Zea mays genomic clone ZMMBta133M02,
  genomic survey sequence.
ACCESSION
  BZ721194
VERSION
  BZ721194.1 GI:28513208
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
    clade; Panicoideae; Andropogoneae; Zea.
    1 (bases 1 to 809)
    Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
    A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
    Maize Genomics Consortium
    Unpublished
    Contact: Cathy Whitelaw
    TIGR
    9712 Medical Center Drive, Rockville, MD 20850, USA
    Tel: 301-838-5843
    Fax: 301-838-0208
    Email: whitelaw@tigr.org
    Seq primer: TF
    Class: sheared ends.
  Location/Qualifiers
    1. .809
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBta133M02"
      /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"
    231 a 173 c 172 g 233 t
    BASE COUNT
    ORIGIN
      Query Match 85.7%; Score 18; DB 29; Length 809;
      Best Local Similarity 100.0%; Pred. No. 5.8e+02;
      Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    Qy 4 CCTGATTGATGACCCAGGG 21
    Db 773 CCTGATTGATGACCCAGGG 790
  
```

```

RESULT 3
AA894450/c
LOCUS
DEFINITION
  aa894450 s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1417861 3',
  mRNA sequence.
ACCESSION
  AA894450
VERSION
  AA894450.1 GI:3030851
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 307)
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
  JOURNAL
    Unpublished
  COMMENT
    Contact: Robert Strausberg, Ph.D.
    Email: gcapbs-remail.nih.gov
    Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
    Emmert-Buck, M.D., Ph.D.
    cDNA Library Preparation: M. Bento Soares, Ph.D.
    cDNA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    www-bio.llnl.gov/bbrp/image/image.html
    Seq primer: -40ml3 fwd. B7 from Amersham
    High quality sequence stop: 288.
  Location/Qualifiers
    1. .307
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:1417861"
      /tissue_type="carcinoid"
      /lab_host="DH10B"
      /clone_lib="NCI_CGAP_Lu5"
      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
      modified polylinker; 1st strand cDNA was prepared from
      neuroendocrine lung carcinoid, and was then primed with a
      Not I - oligo(dT) primer. Double-stranded cDNA was ligated
      to Eco RI adaptors (Pharmacia), digested with Not I and
      cloned into the Not I and Eco RI sites of the modified
      pT7T3 vector. Library is normalized. Library was
      constructed by Bento Soares and M. Fatima Bonaldo. "
    116 a 62 c 42 g 87 t
    BASE COUNT
    ORIGIN
      Query Match 84.8%; Score 17.8; DB 9; Length 307;
      Best Local Similarity 90.5%; Pred. No. 4.6e+02;
      Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    Qy 1 AAGCTGATTGATGACCCAGGG 21
    Db 247 AAACCTGTTGATGACCCAGGG 227

RESULT 4
AI572687/c
LOCUS
DEFINITION
  AI572687 tr78B10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2224411 3',
  mRNA sequence.
ACCESSION
  AI572687
VERSION
  AI572687.1 GI:4536061
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 339)
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
  JOURNAL
    Unpublished
  COMMENT
    Contact: Robert Strausberg, Ph.D.
    Email: gcapbs-remail.nih.gov
    Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
    Emmert-Buck, M.D., Ph.D.
    cDNA Library Preparation: M. Bento Soares, Ph.D.
    cDNA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    www-bio.llnl.gov/bbrp/image/image.html
    Seq primer: -40ml3 fwd. B7 from Amersham
    High quality sequence stop: 288.
  Location/Qualifiers
    1. .307
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:1417861"
      /tissue_type="carcinoid"
      /lab_host="DH10B"
      /clone_lib="NCI_CGAP_Lu5"
      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
      modified polylinker; 1st strand cDNA was prepared from
      neuroendocrine lung carcinoid, and was then primed with a
      Not I - oligo(dT) primer. Double-stranded cDNA was ligated
      to Eco RI adaptors (Pharmacia), digested with Not I and
      cloned into the Not I and Eco RI sites of the modified
      pT7T3 vector. Library is normalized. Library was
      constructed by Bento Soares and M. Fatima Bonaldo. "
    116 a 62 c 42 g 87 t
    BASE COUNT
    ORIGIN
      Query Match 84.8%; Score 17.8; DB 9; Length 307;
      Best Local Similarity 90.5%; Pred. No. 4.6e+02;
      Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    Qy 1 AAGCTGATTGATGACCCAGGG 21
    Db 247 AAACCTGTTGATGACCCAGGG 227

RESULT 4
AI572687/c
LOCUS
DEFINITION
  AI572687 tr78B10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2224411 3',
  mRNA sequence.
ACCESSION
  AI572687
VERSION
  AI572687.1 GI:4536061
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 339)
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
  JOURNAL
    Unpublished
  COMMENT
    Contact: Robert Strausberg, Ph.D.
    Email: gcapbs-remail.nih.gov
    Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
    Emmert-Buck, M.D., Ph.D.
    cDNA Library Preparation: M. Bento Soares, Ph.D.
    cDNA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    www-bio.llnl.gov/bbrp/image/image.html
    Seq primer: -40ml3 fwd. B7 from Amersham
    High quality sequence stop: 288.
  Location/Qualifiers
    1. .307
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:1417861"
      /tissue_type="carcinoid"
      /lab_host="DH10B"
      /clone_lib="NCI_CGAP_Lu5"
      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
      modified polylinker; 1st strand cDNA was prepared from
      neuroendocrine lung carcinoid, and was then primed with a
      Not I - oligo(dT) primer. Double-stranded cDNA was ligated
      to Eco RI adaptors (Pharmacia), digested with Not I and
      cloned into the Not I and Eco RI sites of the modified
      pT7T3 vector. Library is normalized. Library was
      constructed by Bento Soares and M. Fatima Bonaldo. "
    116 a 62 c 42 g 87 t
    BASE COUNT
    ORIGIN
      Query Match 84.8%; Score 17.8; DB 9; Length 307;
      Best Local Similarity 90.5%; Pred. No. 4.6e+02;
      Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    Qy 1 AAGCTGATTGATGACCCAGGG 21
    Db 247 AAACCTGTTGATGACCCAGGG 227
  
```

```

Class: BAC ends.
Location/Qualifiers
1. .380
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SeNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-53J16"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/notes="Vector: PTABAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SeNHsd/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT      105 a      89 c      103 g      83 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 28; Length 380;
Best Local Similarity 90.5%; Pred.No. 5e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2;

QY      1 AAGCCTGATTGATGACCAAGG 21
|||||
Db      147 AAGCCTGATTGATGACAG 167

CA941354      545 bp      mRNA      linear      EST 30-DEC-2002
ir34b01.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6546961 3',
mRNA sequence.
CA941354
CA941354.1 GI:27429834
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 545)
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,B., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagaris,hvll,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 452.
Location/Qualifiers
1. .545
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6546961"
/tissue_type="Purified pancreatic islet"
/lab host="DH10B"
/clone_lib="HR85 islet"
/notes="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site 2: XhoI; cDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'

```

XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."

BASE COUNT 184 a 100 c 88 g 173 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 545;
Best Local Similarity 90.5%; Pred. No. 6e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTTGATTGATGACCAGG 21
|||||
Db 257 AAACCTGGTTGATGACCAGG 237

RESULT 7
AI816423/c

LOCUS
DEFINITION au47g07.xl Schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2517948 3', mRNA sequence.

ACCESSION AI816423

VERSION AI816423.1 GI:5431969

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 592)
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 469.

FEATURES
source
1..592
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2517948"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/clone_lib="Schneider fetal brain 00004"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site 1: S81; Site 2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCCGCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGAGCTCGGTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 Kb inserts and has an average
insert size estimated at 1.2 Kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."

BASE COUNT 195 a 106 c 98 g 192 t 1 others

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 592;
Best Local Similarity 90.5%; Pred. No. 6.2e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTTGATTGATGACCAGG 21
|||||
Db 259 AAACCTGGTTGATGACCAGG 239

RESULT 8
AG121419/c

LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-130H10.F, genomic survey sequence.

ACCESSION AG121419

VERSION AG121419.1 GI:16650584

KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM Pan troglodytes

REFERENCE 1 (bases 1 to 671)
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE BAC end sequences of Library PTB

REFERENCE 2 (bases 1 to 671)
Unpublished

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE Direct Submission

JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@sc.riken.go.jp, URL:http://bgp.gsc.riken.go.jp/,
tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS
Sequencing: -21M13

LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.

FEATURES
source
1..671
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-130H10.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 195 a 161 c 152 g 163 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 29; Length 671;
Best Local Similarity 90.5%; Pred. No. 6.6e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTTGATTGATGACCAGG 21
|||||
Db 414 AAGCTTGATTGATGACCAGG 394

RESULT 9
AG186976/c

LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-004E01.R, genomic survey sequence.

ACCESSION AG186976

VERSION AG186976.1 GI:18149509

KEYWORDS GSS.

SOURCE
ORGANISM Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H., and Sakaki, Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H., and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suhiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbesgsc.riken.go.jp, URL: http://hgp-gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
LOCATION/Qualifiers
1. 708
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-004E01.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 201 a 165 c 156 g 182 t 3 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 29; Length 708;
Best Local Similarity 90.5%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTCATGACCCAGG 21
|||||
Db 398 AAGCTGATTCATGACCCAGG 378
|||||

RESULT 10
LOCUS BU245702 913 bp mRNA linear EST 26-NOV-2002
DEFINITION 603783316F1 CSEQCHN34 Gallus gallus CDNA clone CHEST73504 5', mRNA
sequence.
ACCESSION BU245702
VERSION BU245702.1 GI:25493069
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 913)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1. 913
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST73504"
/sex="Female"
/dev stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN34"
/note="Organ: liver; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

BASE COUNT 217 a 230 c 207 g 259 t
ORIGIN

Query Match 84.8%; Score 17.8; DB 13; Length 913;
Best Local Similarity 90.5%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTCATGACCCAGG 21
|||||
Db 145 AAGCTGATTCATGACCCAGG 165
|||||

RESULT 11
LOCUS BG119393 963 bp mRNA linear EST 30-JAN-2001
DEFINITION 602343076F1 NIH_MGC_90 Homo sapiens CDNA clone IMAGE:4444074 5',
mRNA sequence.
ACCESSION BG119393
VERSION BG119393.1 GI:12612899
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10219 row: 9 column: 19
High quality sequence stop: 658.
LOCATION/Qualifiers
1. 963
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4444074"

```

/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 90"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH MGC Library."
BASE COUNT      285 a 199 c 199 g 280 t
ORIGIN

Query Match      84.8%; Score 17.8; DB 10; Length 963;
Best Local Similarity 90.5%; Pred. No. 7.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATGACCCAGG 21
    |||||
Db 570 AAACCTGGTTGATGACCCAGG 590

RESULT 12
AL523235/c
LOCUS      1201 bp      mRNA      linear      EST 22-MAY-2003
DEFINITION Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CSODC001YE18 3-PRIME, mRNA sequence.
ACCESSION  AL523235
VERSION     AL523235.2 GI:31041496
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 1201)
AUTHORS   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished
COMMENT   On Feb 13, 2001 this sequence version replaced gi:12786728.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10668.r, and
it belongs to a clone representative of this cluster. For more
information about this cluster and the virtual cDNA, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODC001BC09NP1&cluster=10668.r. Contact :
Feng Liang Email : fliang@lifetech.com URL : Corporation 1600
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODC001BC09NP1.
FEATURES
source
1..1201
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODC001YE18"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      358 a 206 c 252 g 311 t 74 others
ORIGIN

Query Match      84.8%; Score 17.8; DB 9; Length 1201;
Best Local Similarity 90.5%; Pred. No. 8.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATGACCCAGG 21
    |||||
Db 236 AAACCTGGTTGATGACCCAGG 216

/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 90"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH MGC Library."
BASE COUNT      285 a 199 c 199 g 280 t
ORIGIN

```

```

RESULT 13
AZ072191
LOCUS      565 bp      DNA      linear      GSS 31-MAR-2000
DEFINITION RPCI-23-39618.TV RPCI-23 Mus musculus genomic clone RPCI-23-39618,
genomic survey sequence.
ACCESSION  AZ072191
VERSION    AZ072191.1 GI:7365088
KEYWORDS   GSS.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 565)
AUTHORS   Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
TITLE     Mouse BAC End Sequences from Library RPCI-23
JOURNAL   Unpublished
COMMENT   Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 396 row: I column: 8
Seq primer: T7
Class: BAC ends.
FEATURES
source
1..565
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-39618"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      142 a 131 c 162 g 130 t
ORIGIN

Query Match      82.9%; Score 17.4; DB 28; Length 565;
Best Local Similarity 94.7%; Pred. No. 9.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCTGATTGATGACCCAGG 20
    |||||
Db 26 AGCCTGATTGATGACCCAGG 44

Query Match      82.9%; Score 17.4; DB 28; Length 565;
Best Local Similarity 94.7%; Pred. No. 9.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCTGATTGATGACCCAGG 20
    |||||
Db 26 AGCCTGATTGATGACCCAGG 44

RESULT 14
BE729033/c
LOCUS      914 bp      mRNA      linear      EST 15-SEP-2000
DEFINITION 601561212F1 NIH MGC 20 Homo sapiens cDNA clone IMAGE:3830835 5',
mRNA sequence.
ACCESSION  BE729033
VERSION    BE729033.1 GI:10143025
KEYWORDS   EST.
SOURCE     Homo sapiens (human)

```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 914)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCW505 row: h column: 04
 High quality sequence stop: 694.

FEATURES

source
 1..914
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3830835"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 20"
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 212 a 241 c 238 g 223 t
 ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 914;
 Best Local Similarity 94.7%; Pred. No. 1.2e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATGACCAG 19
 |||||
 Db 626 AAGCTGATTGATGACCAG 608

RESULT 15
 AT818091/c
 LOCUS 543 bp mRNA linear EST 21-DEC-1999
 DEFINITION WK27a09.x1 NCI CGAP Brn25 Homo sapiens cDNA clone IMAGE:2413528 3',
 similar to SW:HS1_FACSU 034520 ATP PHOSPHORIBOSYLTRANSFERASE 1,
 mRNA sequence.

ACCESSION AT818091
 VERSION AT818091.1 GI:5437170
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 543)
 AUTHORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 (CGAP/BTGAP), Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
 Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 904 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 485.
 Location/Qualifiers

FEATURES

source
 1..543
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2413528"
 /tissue_type="anaplastic oligodendroglioma"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Brn25"
 /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT7T3 vector.
 Library is normalized, and was constructed by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 99 a 174 c 180 g 90 t
 ORIGIN

Query Match 81.0%; Score 17; DB 9; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCCTGATTGATGACCAG 19
 |||||
 Db 494 GCCTGATTGATGACCAG 478

Search completed: February 19, 2004, 23:14:36
 Job time : 1575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 19:59:22 ; Search time 175.5 Seconds
(without alignments)
323.010 Million cell updates/sec

Title: US-10-085-108-21_COPY_711_731

Perfect score: 21

Sequence: 1 AAGCCTGATTGATGACCAGG 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 19Jun03.*

1: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2002.DAT.*
25: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21	100.0	21	25	Human MAGE-C3 expr
C 2	21	100.0	425	22	Human breast cell
C 3	21	100.0	425	22	Human foetal liver
C 4	21	100.0	425	22	Probe #4696 for ge
C 5	21	100.0	425	22	Human brain expres
C 6	21	100.0	425	22	Human bone marrow
C 7	21	100.0	425	22	Probe #4815 for ge
C 8	21	100.0	425	22	Probe #4923 used t

C 9	21	100.0	425	22	AAI04666	Probe #4657 used t
C 10	21	100.0	425	23	ABS29927	Human liver single
C 11	21	100.0	425	24	ABS04866	Human genome-deriv
C 12	21	100.0	1041	25	ABX95006	CNA encoding huma
C 13	21	100.0	7806	16	ABX88354	DNA encoding novel
C 14	17	81.0	336	16	AAI24776	Human gene signatu
C 15	16.8	80.0	504	24	ABS69805	Novel murine polyn
C 16	16.2	77.1	202	19	AAI10827	Human biallelic po
C 17	16.2	77.1	419	23	ABV17420	Human prostate exp
C 18	16.2	77.1	456	22	ABA58928	Human foetal liver
C 19	16.2	77.1	456	22	AAK07090	Human brain expres
C 20	16.2	77.1	456	22	AAK32830	Human bone marrow
C 21	16.2	77.1	456	22	AAI38645	Probe #7331 used t
C 22	16.2	77.1	456	23	ABS32555	Human liver single
C 23	16.2	77.1	456	24	ABS07634	Human genome-deriv
C 24	16.2	77.1	458	23	ABV47214	Human prostate exp
C 25	16.2	77.1	964	22	ABA15191	Human nervous syst
C 26	16.2	77.1	1236	23	ABL03601	Drosophila melanog
C 27	16.2	77.1	1781	22	ABAI8978	Human nervous syst
C 28	16.2	77.1	1920	22	ABAI0827	Manihot esculenta
C 29	16.2	77.1	2204	22	ABA15194	Human nervous syst
C 30	16.2	77.1	3171	22	AAK85455	Cold-active beta g
C 31	16.2	77.1	3479	23	ABL03600	Drosophila melanog
C 32	16.2	77.1	3968	23	ABL03602	Drosophila melanog
C 33	16.2	77.1	4912	21	AAA65349	NDO related comple
C 34	16.2	77.1	6779	21	AAA65350	NDO related comple
C 35	16.2	77.1	7325	24	AA894817	Human DNA sequence
C 36	15.8	75.2	470	22	ABA26707	Probe #5173 for ge
C 37	15.8	75.2	1378	23	ABL07577	Drosophila melanog
C 38	15.8	75.2	3116	23	ABL03310	Drosophila melanog
C 39	15.8	75.2	3378	23	ABL07576	Drosophila melanog
C 40	15.8	75.2	4764	23	ABL05675	Drosophila melanog
C 41	15.8	75.2	6147	22	ABA14804	Human nervous syst
C 42	15.8	75.2	7300	23	ABL05674	Drosophila melanog
C 43	15.4	73.3	312	22	AAH66909	C glutamic codin
C 44	15.4	73.3	320	22	AAK87704	Human digestive sy
C 45	15.4	73.3	320	22	AA831687	Human liver associ

ALIGNMENTS

RESULT 1

ABX95022/c

ID ABX95022 standard; DNA; 21 BP.

XX ABX95022;

AC ABX95022;

XX 05-JUN-2003 (first entry)

DT 05-JUN-2003 (first entry)

XX Human MAGE-C3 expression pattern anlaysis RT-PCR antisense primer.

DE TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;

XX tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;

KW head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;

KW cutaneous melanoma; nonsmall cell lung cancer; RT-PCR; primer; MAGE-C3;

KW human; reverse transcription.

XX Homo sapiens.

OS US2002176865-A1.

FN 28-NOV-2002.

XX 01-MAR-2002; 2002US-0085108.

PF 09-FEB-2000; 2000US-0501104.

PR 25-APR-1997; 97US-0845528.

PR 24-APR-1998; 98US-0066281.

PR 17-DEC-1999; 99US-0468433.

XX (LUCAS/)

PA (BOON/)

BOON-FALLEUR T.

XX Lucas S, Boon-Falleur T;
 XX WPI; 2003-328468/31.
 XX
 XX Novel isolated nucleic acid encoding tumor rejection antigen precursor
 PT MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
 PT presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
 PT MAGE-B6
 XX
 XX Example 12; Page 13; 59pp; English.
 XX
 XX The invention relates to an isolated nucleic acid molecule which encodes
 CC a tumour rejection antigen precursor (TRAP) having an amino acid sequence
 CC of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
 CC polynucleotide sequence. Also disclosed is a method which is useful for
 CC determining presence of cytolytic T-cells specific for complexes of human
 CC leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
 CC cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
 CC useful as a diagnostic probe to determine the presence of abnormal
 CC (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
 CC head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
 CC cutaneous melanoma or non-small cell lung cancer (NSCLC) which express
 CC MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
 CC disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
 CC or tumour rejection antigens (TRAs). The present sequence represents the
 CC human MAGE-C3 expression pattern analysis reverse transcription (RT)-PCR
 CC antisense primer.
 XX
 XX Sequence 21 BP; 4 A; 7 C; 4 G; 6 T; 0 other;
 SQ
 Query Match 100.0%; Score 21; DB 25; Length 21;
 Best Local Similarity 100.0%; Pred. No. 0.39;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGCCTGATTGATGACCAGG 21
 Db 21 AAGCCTGATTGATGACCAGG 1
 RESULT 2
 ABA46075/c
 ID ABA46075 standard; DNA; 425 BP.
 XX
 AC ABA46075;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #4770.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for

DR WPI; 2001-496933/54.
 XX
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 XX Claim 1; SEQ ID NO 4770; 327pp + sequence listing; English.
 XX
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;
 SQ
 Query Match 100.0%; Score 21; DB 22; Length 425;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGCCTGATTGATGACCAGG 21
 Db 150 AAGCCTGATTGATGACCAGG 130
 RESULT 3
 ABA56617/c
 ID ABA56617 standard; DNA; 425 BP.
 XX
 AC ABA56617;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #4922.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for

Db 150 AAGCCTGATTGATGACCAGG 130
|||||

RESULT 6

AAK30282/c
ID AAK30282 standard; DNA; 425 BP.

XX AC AAK30282;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed single exon probe SEQ ID NO: 4839.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00668.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US-024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PD WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -

XX PS Example 4; SEQ ID NO: 4839; 658pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX CC the probes of the invention.

XX SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. NO. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCCTGATTGATGACCAGG 21
|||||

Db 150 AAGCCTGATTGATGACCAGG 130
|||||

RESULT 7

AAI14882/c

ID AAI14882 standard; DNA; 425 BP.

XX AC AAI14882;

XX DT 12-OCT-2001 (first entry)

XX DE Probe #4815 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;

KW cervical cancer; ss.

OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US-024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PD WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -

XX PS Claim 25; SEQ ID NO 4815; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENP). The present sequence is one such probe. The SENPs are derived
XX CC from human Hela cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.

XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. NO. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCCTGATTGATGACCAGG 21
|||||

Db 150 AAGCCTGATTGATGACCAGG 130
|||||

RESULT 8

AAI36237/c

ID AAI36237 standard; DNA; 425 BP.

XX AC AAI36237;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #4923 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 4923; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;
Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTGATTGATGACACGAGG 21
DB 150 AAGCTGATTGATGACACGAGG 130
RESULT 9
AAI04666/c
ID AAI04666 standard; DNA; 425 BP.
XX
XX AAI04666;
AC
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #4657 used to measure gene expression in human breast sample.
XX
XX Probe; human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX W0200157270-A2.
PN
XX 09-AUG-2001.
PD
XX 29-JAN-2001; 2001WO-US00661.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX

PS Claim 25; SEQ ID No 4657; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;
Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTGATTGATGACACGAGG 21
DB 150 AAGCTGATTGATGACACGAGG 130
RESULT 10
ABS29927/c
ID ABS29927 standard; DNA; 425 BP.
XX
XX ABS29927;
AC
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID No 4917.
XX
XX Human; single exon nucleic acid probe; liver; cirrhosis;
XX hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
XX coronary heart disease; ss.
XX
XX Homo sapiens.
OS
XX W0200157273-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00664.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human adult liver -
XX
XX Claim 1; SEQ ID No 4917; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (1) may be used for predicting, measuring and displaying gene
 CC expression in samples derived from human adult liver. The genes
 CC identified may be involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
 CC is associated with coronary heart disease. AB525011-AB551005 represent
 CC human liver single exon nucleic acid probes of the invention.
 CC Note: The sequence information for this patent does not appear in the
 CC printed specification but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 23; Length 425;

Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCCTGATTGATGACGAGG 21

DB 150 AAGCCTGATTGATGACGAGG 130

RESULT 11

AB504866/c

ID ABS04866 standard; DNA; 425 BP.

XX AC ABS04866;

XX DT 19-AUG-2002 (first entry)

XX DE Human genome-derived single exon probe from lung SEQ ID No 4857.

XX KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;

XX KW Chronic obstructive pulmonary disease; interstitial lung disease;

XX KW Familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX KW Tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;

XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;

XX KW Pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX KW Pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX KW Primary ciliary dyskinesia; pulmonary hypertension;

XX KW Hyaline membrane disease.

XX OS Homo sapiens.

XX XX WO200186003-A2.

XX XX 15-NOV-2001.

XX XX 30-JAN-2001; 2001WO-US00665.

XX XX 04-FEB-2000; 2000US-180312P.

XX XX 26-MAY-2000; 2000US-207456P.

XX XX 30-JUN-2000; 2000US-0608408.

XX XX 03-AUG-2000; 2000US-0632366.

XX XX 21-SEP-2000; 2000US-234687P.

XX XX 27-SEP-2000; 2000US-236359P.

XX XX 04-OCT-2000; 2000US-0024263.

XX XX (MOLE-) MOLECULAR DYNAMICS INC.

XX XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2002-114193/15.

XX XX Spatially-addressable set of single exon nucleic acid probes, used to

XX XX measure gene expression in human lung samples -

XX XX Claim 1; SEQ ID No 4857; 634pp; English.

XX XX The invention relates to a spatially-addressable set of single exon

XX XX nucleic acid probes for measuring gene expression in a sample derived

XX XX from human lung comprising single exon nucleic acid probes having one of

XX XX 12614 nucleic acid sequences mentioned in the specification, or their

XX XX complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberosus sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic,
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe of the invention.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 24; Length 425;

Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCCTGATTGATGACGAGG 21

DB 150 AAGCCTGATTGATGACGAGG 130

RESULT 12

ABX95006

ID ABX95006 standard; cDNA; 1041 BP.

XX AC ABX95006;

XX DT 05-JUN-2003 (first entry)

XX DE cDNA encoding human tumour rejection antigen precursor, MAGE-C3.

XX KW TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;

XX KW tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;

XX KW head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;

XX KW cutaneous melanoma; non-small cell lung cancer; gene; MAGE-C3; human;

XX KW chromosome Xq27.1-Xq27.3.

XX OS Homo sapiens.

XX XX

XX FH Key Location/Qualifiers

XX FT CDS 1..1041

XX FT /*tag= a

XX FT /product= "MAGE-C3"

XX XX

XX PN US2002176865-A1.

XX XX

XX XX 28-NOV-2002.

XX PF 01-MAR-2002; 2002US-0085108.
 XX PR 09-FEB-2000; 2000US-0501104.
 XX PR 25-APR-1997; 97US-0845528.
 XX PR 24-APR-1998; 98US-0066281.
 XX PR 17-DEC-1999; 99US-0468433.
 XX PA (LUCAS/) LUCAS S.
 XX PA (BOON/) BOON-FALLEUR T.
 XX PI Lucas S, Boon-Falleur T;
 XX DR WPI; 2003-328468/31.
 XX DR P-PSDB; ABU08932.
 XX PT Novel isolated nucleic acid encoding tumor rejection antigen precursor
 XX PT MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
 XX PT presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
 XX PT MAGE-B6 -
 XX PS Claim 1; Fig 5; 59pp; English.
 XX CC The invention relates to an isolated nucleic acid molecule which encodes
 XX CC a tumour rejection antigen precursor (TRAP) having an amino acid sequence
 XX CC of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
 XX CC polynucleotide sequence. Also disclosed is a method which is useful for
 XX CC determining presence of cytolytic T-cells specific for complexes of human
 XX CC leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
 XX CC cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
 XX CC useful as a diagnostic probe to determine the presence of abnormal
 XX CC (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
 XX CC head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
 XX CC cutaneous melanoma or non-small cell lung cancer (NSCLC) which express
 XX CC MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
 XX CC disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPS
 XX CC or tumour rejection antigens (TRAs). The present sequence represents the
 XX CC cDNA of the gene encoding the human tumour rejection antigen precursor,
 XX CC MAGE-C3, which is located on chromosome Xq27.1-Xq27.3.
 XX SQ Sequence 1041 BP; 242 A; 283 C; 242 G; 274 T; 0 other;
 Query Match 100.0%; Score 21; DB 25; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.81; Mismatches 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGCCTGATTGATGACCGG 21
 DB 711 AAGCCTGATTGATGACCGG 731
 RESULT 13
 ID AAS88354
 ID AAS88354 standard; cDNA; 7806 BP.
 AC AAS88354;
 XX DT 13-FEB-2002 (first entry)
 XX DE DNA encoding novel human diagnostic protein #24158.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR P-PSDB; ABG24167.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX PT diagnostics, forensics, gene mapping, identification of mutations
 XX PT responsible for genetic disorders or other traits and to assess
 XX PT biodiversity -
 XX PS Claim 1; SEQ ID No 24158; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX CC and gene mapping, and in recombinant production of (II). The
 XX CC polynucleotides are also used in diagnostics as expressed sequence tags
 XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
 XX CC to restore normal activity of (II) or to treat disease states involving
 XX CC (II). (II) is useful for generating antibodies against it, detecting or
 XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
 XX CC a food supplement. (II) and its binding partners are useful in medical
 XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
 XX CC disorders involving aberrant protein expression or biological activity.
 XX CC The polypeptide and polynucleotide sequences have applications in
 XX CC diagnostics, forensics, gene mapping, identification of mutations
 XX CC responsible for genetic disorders or other traits to assess biodiversity
 XX CC and to produce other types of data and products dependent on DNA and
 XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
 XX CC diagnostic coding sequences of the invention.
 XX CC Note: The sequence data for this patent did not appear in the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 7806 BP; 1755 A; 2185 C; 1760 G; 2106 T; 0 other;
 Query Match 100.0%; Score 21; DB 23; Length 7806;
 Best Local Similarity 100.0%; Pred. No. 1.2; Mismatches 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGCCTGATTGATGACCGG 21
 DB 3971 AAGCCTGATTGATGACCGG 3991
 RESULT 14
 AAT24776
 ID AAT24776 standard; cDNA to mRNA; 336 BP.
 XX AC AAT24776;
 XX DT 09-OCT-1996 (first entry)
 XX DE Human gene signature HUMGS06854.
 XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 XX KW human; cloning; mapping; non-biased library; diagnosis; detection;
 XX KW cell typing; abnormal cell function; ss.
 XX OS Homo sapiens.
 XX PN WO9514772-A1.
 XX XX 01-JUN-1995.
 XX PF 11-NOV-1994; 94WO-JP01916.
 XX PR 12-NOV-1993; 93JP-0355504.

PA (MATS/) MATSUBARA K.
 XX (OKUB/) OKUBO K.
 XX Matsubara K, Okubo K;
 PI WPI; 1995-206931/27.
 XX
 DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 XX Claim 1; Page 1692; 2245pp; Japanese.
 PS
 XX A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 XX
 SQ Sequence 336 BP; 91 A; 67 C; 80 G; 83 T; 15 other;
 Query Match 81.0%; Score 17; DB 16; Length 336;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 CTGATTGATGACACGGG 21
 Db 231 CTGATTGATGACACGGG 247
 |||||
 RESULT 15
 ABS69805
 ID ABS69805 standard; DNA; 504 BP.
 XX
 AC ABS69805;
 XX
 DT 21-NOV-2002 (first entry)
 XX
 DE Novel murine polynucleotide isolated using gene trap technology #869.
 XX
 KW Mouse; gene trapped sequence; GTS; functional genomic analysis;
 KW phage display system; gene chip; temporal gene expression;
 KW tissue specific gene expression; antisense inhibition; gene targeting;
 KW development disorder; cell differentiation disorder; aging; cancer;
 KW autoimmune disease; lupus; inflammatory disorder; skin disorder;
 KW degenerative disorder; ds.
 XX
 OS Mus musculus.
 XX
 XX US2002102543-A1.
 PN
 XX 01-AUG-2002.
 PD
 XX 30-NOV-2000; 2000US-0728445.
 PF
 XX 01-DEC-1999; 99US-168358P.
 PR
 XX (FRIE/) FRIEDRICH G.
 PA (ZAMB/) ZAMBROWICZ B.
 XX (SAND/) SANDS A T.
 PI Friedrich G, Zambrowicz B, Sands AT;

XX WPI; 2002-690598/74.
 DR
 XX Novel murine polynucleotides that individually identify novel genes
 PT into which a retroviral gene trap vector has integrated, useful in
 PT genomic analysis and in discovery, development of therapeutic and
 PT diagnostic agents -
 XX
 XX Claim 1; Page 286; 296pp; English.
 PS
 XX The invention describes an isolated murine polynucleotide (I) comprising
 CC a contiguous stretch of at least 60 nucleotides of one of 265-677
 CC nucleotide 891 OMNIBANK gene trapped sequences (GTSs) (S), given in the
 CC specification. The novel genes and cells are useful in functional
 CC genomic analysis and in the discovery and development of new therapeutic
 CC and diagnostic agents and methods. (I) is useful for identifying the
 CC coding regions of the murine genome, to isolate cDNAs, genomic clones,
 CC or full-length genes/polynucleotides or homologues, heterologues,
 CC paralogues, or orthologues that are capable of hybridising to one or more
 CC of the GTSs under stringent conditions. (I) can be incorporated into a
 CC phage display system that can be used to screen for proteins or other
 CC ligands, that are capable of binding an amino acid sequence encoded by
 CC an oligonucleotide or polynucleotide sequence in at least one of the TS
 CC sequences. (I) is useful in addressable arrays, such as gene chips, to
 CC identify and characterise temporal and tissue specific gene expression,
 CC to identify the gene of interest from many sources and for genetic
 CC manipulations such as antisense inhibition and gene targeting. Decreasing
 CC the level of expression of (I) and/or down regulating the activity of
 CC peptides or proteins encoded by (I) is useful for treating development
 CC and cell differentiation disorders, aging, cancer, autoimmune disease,
 CC lupus, inflammatory disorders, skin disorders and degenerative
 CC disorders. This sequence represents a murine cDNA isolated using gene
 CC trap technology.
 XX
 SQ Sequence 504 BP; 110 A; 152 C; 146 G; 96 T; 0 other;
 Query Match 80.0%; Score 16.8; DB 24; Length 504;
 Best Local Similarity 90.0%; Pred. No. 1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 AGCCTGATTGATGACACGGG 21
 Db 461 AGCAGGATTGATGACACGGG 480
 |||||
 Search completed: February 19, 2004, 21:28:33
 Job time : 176.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 20:01:08 ; Search time 1596.5 Seconds
(without alignments)
538.116 Million cell updates/sec

Title: US-10-085-108-21_COPY_711_731
Perfect score: 21
Sequence: 1 AAGCCTGATTGATGACCAGGG 21

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	21	100.0	1932	9	AF490508	AF490508 Homo sapi
2	21	100.0	124497	9	HS232G24	AL022152 Human DNA
C 3	19	90.5	153439	2	EX323047	BX323047 Danio rer
C 4	19	90.5	159948	2	BX248240	BX248240 Danio rer
5	17.8	84.8	1494	9	AK098443	AK098443 Homo sapi
6	17.8	84.8	1681	9	BC048268	BC048268 Homo sapi
C 7	17.8	84.8	67201	9	AL136169	AL136169 Human DNA
8	17.8	84.8	163746	2	AC102710	AC102710 Mus muscu
C 9	17.8	84.8	174191	9	AC109462	AC109462 Homo sapi
C 10	17.8	84.8	177733	2	AC113908	AC113908 Rattus no
C 11	17.8	84.8	180353	2	AC128288	AC128288 Rattus no
C 12	17.8	84.8	186118	2	AC117856	AC117856 Rattus no
C 13	17.8	84.8	194454	2	AC111040	AC111040 Mus muscu
C 14	17.8	84.8	194520	10	AL691424	AL691424 Mouse DNA
C 15	17.8	84.8	194636	10	AC123058	AC123058 Mus muscu
C 16	17.8	84.8	214493	2	AC097047	AC097047 Rattus no
C 17	17.8	84.8	227369	2	AC114469	AC114469 Rattus no
C 18	17.8	84.8	231827	2	AC096130	AC096130 Rattus no
C 19	17.8	84.8	232476	2	AC18621	AC18621 Mus muscu
C 20	17.8	84.8	243257	2	AC130512	AC130512 Rattus no
21	17.8	84.8	251759	2	AC137283	AC137283 Rattus no
C 22	17.8	84.8	257104	2	AC122547	AC122547 Mus muscu
C 23	17.8	84.8	274024	2	AC102995	AC102995 Rattus no
C 24	17.8	84.8	274467	2	AC113707	AC113707 Rattus no
C 25	17.4	82.9	196	5	AY185781	AY185781 Ambystoma
C 26	17.4	82.9	197	5	AY185782	AY185782 Ambystoma
C 27	17.4	82.9	223	5	AY185785	AY185785 Ambystoma
C 28	17.4	82.9	223	5	AY185786	AY185786 Ambystoma
C 29	17.4	82.9	224	5	AY185783	AY185783 Ambystoma
C 30	17.4	82.9	224	5	AY185787	AY185787 Ambystoma
C 31	17.4	82.9	227	5	AY185784	AY185784 Ambystoma
C 32	17.4	82.9	4202	5	AF001958	AF001958 Ambystoma
C 33	17.4	82.9	92337	10	AL604022	AL604022 Mouse DNA
C 34	17.4	82.9	120492	3	CEY43F8C	AL032637 Caenorhab
C 35	17.4	82.9	198727	9	AC122401	AC122401 Mus muscu
C 36	17.4	82.9	20361	9	AC011374	AC011374 Homo sapi
C 37	17.4	82.9	218092	2	AC019272	AC019272 Mus muscu
C 38	17.4	82.9	219451	2	AC137244	AC137244 Rattus no
39	17.4	82.9	226771	2	AC110940	AC110940 Rattus no
40	17.4	82.9	242602	2	AC095427	AC095427 Rattus no
41	17	81.0	108542	10	AL627307	AL627307 Mouse DNA
42	17	81.0	208966	2	AC099588	AC099588 Mus muscu
C 43	17	81.0	212119	2	AC099602	AC099602 Mus muscu
C 44	17	81.0	236604	2	AC099619	AC099619 Mus muscu
C 45	16.8	80.0	403	9	AF490961	AF490961 Homo sapi

ALIGNMENTS

RESULT 1
AF490508
LOCUS
DEFINITION Homo sapiens hepatocellular carcinoma-associated protein HCA2 mRNA,
complete cds.
ACCESSION AF490508
VERSION AF490508.1 GI:19919741
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1932)
AUTHORS Dong,X. and Chen,W.
TITLE Identification of genes in the chromosome X that are differentially

expressed in hepatocellular carcinoma

Unpublished
2 (bases 1 to 1932)
Dong, X. and Chen, W.
Direct Submission
Submitted (06-MAR-2002) Department of Immunology, T Cell Research
Lab., Xueuan Road 38, Beijing 100083, China

Location/Qualifiers
1..1932
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="X"
1..1932
/codon_start=1
/product="hepatocellular carcinoma-associated protein
HCA2"
/protein_id="AA008355.1"
/db_xref="GI:19919742"
/translation="MLLPCHWVLTATSDSLGOWKNTCATYALSPVLPQPQPRK
KATDKYSAPHLGLREVRFLRGSTSDQRMDSLVLCPTVFKLWRTLSGPGLSLDL
HFGSPQPKTSRLRAVSQREEDPWLNEKRLWKDSLDLPTWRGTGVTLSLPVLS
PKRLWKEKAGSLPFLTYTLDEKVLQVLLKYOAKELTAEQMNVINY
TGYPMTFRKAREFIEILGLSLTEVDPHYVFNLDLTCEGSLDEQMPQNRLL
LLILSVFIKNCASSEVINELNAGFWALGADVLKRLALWSEGEAFCEESG
LRAGSVLDLANPQGLAGHQEDRGRLTEASPOQKGGEDDMPAAGPPLPQSP
EIPQGPPIKSPQGPQSPQSPDSCSPLLTRLDSESSSEEDATWALPES
LPRALDEKVAELVQLLYKQTPVTKAEMLTIVIKKYKQVPMIFGKAHEFIELI
FGIALTMDPDNHSYFFEDTLDLTIEGSLIDDOQMPKNCILLILIFIKGSCVPEE
VIVEVLSAIGPIORPAREVLEFLSKLSIIPSPFWSYMDALMEDRAQIDTTDD
ATMASASPSPVMSNFCPE"

BASE COUNT 477 a 489 c 510 g 456 t

ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 1932;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATCACCAGG 21
|||||

Db 1605 AAGCTGATTGATCACCAGG 1625
|||||

RESULT 2
HS232G24 124497 bp DNA linear PRI 09-MAR-2002
LOCUS
DEFINITION
Human DNA sequence from clone RP6-232G24 on chromosome Xq27.1-27.3
Contains the gene for the melanoma antigen gene family protein,
MAGEC3 and the MAGEC1 gene for melanoma antigen, family C.1,
complete sequence.

ACCESSION
AL022152
AL022152.1 GI:3150086
HTG; MAGEC1; MAGEC3.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 124497)
Direct Submission
Submitted (08-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
Clon request: clonrequest@sanger.ac.uk
On May 22, 1998 this sequence version replaced gi:2969932.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone of more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP6-232G24 is from the library RPCI-6 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pPAC4
This sequence is the entire insert of clone RP6-232G24 The true
right end of clone RP3-326L12 is at 82755 in this sequence.

FEATURES
Location/Qualifiers
source
1..124497
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCI:RP709G24232"
/db_xref="taxon:9606"
/chromosome="X"
/map="q27.1-27.3"
/clone="RP6-232G24"
/clone_lib="RPCI-6"
6..334
/notes="LMC2 repeat: matches 5195..5546 of consensus"
356..762
/notes="LMC2 repeat: matches 5843..6246 of consensus"
763..794
/note="16 copies 2 mer tg 93% conserved"
946..1161
/notes="LMC1 repeat: matches 5354..5569 of consensus"
1181..1399
/notes="LMC1 repeat: matches 5914..6136 of consensus"
1471..1606
/note="L2 repeat: matches 2574..2708 of consensus"
1939..2330
/notes="14 copies 28 mer 57% conserved"
1969..2326
/notes="179 copies 2 mer gg 57% conserved"
1972..2331
/note="20 copies 18 mer 59% conserved"
1996..2265
/notes="5 copies 54 mer 65% conserved"
2029..2181
/notes="3 copies 51 mer 74% conserved"
2153..2347
/note="5 copies 39 mer 65% conserved"
2426..4801
/gene="da232G24.1"
join(2426..2597,3291..3384,3693..4801)
/gene="da232G24.1"
/product="da232G24.1 (Melanoma antigen gene family
protein, MAGEC3)"
/notes="(possible pseudogene)
match: cDNAs: Em:AF151378
match: ESTs: Em:BI460078"
/evidence=not_experimental
2827..2872
/note="23 copies 2 mer gg 78% conserved"
3761..4801
/gene="da232G24.1"
/note="(possible pseudogene)
match: proteins: Sw:P43363 Tr:Q9BZ80 Tr:O60781 Tr:AAH04105
Tr:O95529 Tr:Q9R2A2 Sw:O15479 Tr:O89006 Tr:Q9BUN9
Tr:Q96M61"
/codon_start=1
/evidence=not_experimental

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3298: contig of 3298 bp in length
* 3299: gap of 100 bp
* 3398: contig of 7210 bp in length
* 10609: gap of 100 bp
* 10709: contig of 27115 bp in length
* 37823: gap of 100 bp
* 37924: contig of 7452 bp in length
* 45375: gap of 100 bp
* 45376: gap of 100 bp
* 45476: contig of 13255 bp in length
* 58730: gap of 100 bp
* 58830: contig of 2396 bp in length
* 61226: gap of 100 bp
* 61327: contig of 3155 bp in length
* 64481: gap of 100 bp
* 64582: contig of 19455 bp in length
* 84036: gap of 100 bp
* 84136: contig of 5025 bp in length
* 89161: gap of 100 bp
* 89261: contig of 24645 bp in length
* 113906: gap of 100 bp
* 114007: contig of 28013 bp in length
* 142019: gap of 100 bp
* 142120: contig of 7156 bp in length
* 149275: gap of 100 bp
* 149276: contig of 4064 bp in length.
* 149376

FEATURES

source

misc_feature 1..3298
/note="assembly_fragment:00374
fragment_chain:1"
3399..10608
/note="assembly_fragment:00202
fragment_chain:1"
10709..37823
/note="assembly_fragment:00149.0"
37924..45375
/note="assembly_fragment:00149.1"
45476..58730
/note="assembly_fragment:00440.0"
58831..61226
/note="assembly_fragment:00440.1"
61327..64481
/note="assembly_fragment:00440.2"
64582..84036
/note="assembly_fragment:00440.3"
84137..89161
/note="assembly_fragment:00707.0"
89262..113906
/note="assembly_fragment:00707.1"
114007..142019
/note="assembly_fragment:00707.2"
142120..149275
/note="assembly_fragment:00861.0"
149376..153439
/note="assembly_fragment:00861.1"
BASE COUNT 32783 a 43384 c 43236 g 32836 t 1200 others
ORIGIN

Query Match

90.5%; Score 19; DB 2; Length 153439;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 GCCTGATTGATGACACGGG 21
|||||
Db 138696 GCCTGATTGATGACACGGG 138678

RESULT 4

LOCUS BX248240/c 159948 bp DNA linear HTG 03-APR-2003
DEFINITION Danio rerio clone CH211-162119, *** SEQUENCING IN PROGRESS ***, 8
unordered pieces.
ACCESSION BX248240
VERSION BX248240.3 GI:29539163
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 159948)
McLaren,S.
AUTHORS Direct Submission
TITLE Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 3, 2003 this sequence version replaced gi:28268576.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZC162119
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 156495 bases at least Q40
Consensus quality: 157186 bases at least Q30
Consensus quality: 158113 bases at least Q20
Insert size: 159248; sum-of-contigs
Quality coverage: 9.26x in Q20 bases; agarose-fp
coverage: 9.80x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3108: contig of 3108 bp in length
* 3109: gap of 100 bp
* 3209: contig of 29682 bp in length
* 32891: gap of 100 bp
* 32991: contig of 16832 bp in length
* 49822: gap of 100 bp
* 49923: contig of 5141 bp in length
* 55064: gap of 100 bp
* 55164: contig of 56681 bp in length
* 111845: gap of 100 bp
* 111945: contig of 19475 bp in length
* 131419: gap of 100 bp
* 131520: contig of 17511 bp in length
* 149030: gap of 100 bp
* 149130: contig of 10818 bp in length.
* 149131: 159948: contig of 10818 bp in length.
Location/Qualifiers
1..159948
/organism="Danio rerio"
/mol_type="genomic DNA"

FEATURES
source

```

/db_xref="taxon:7955"
/clone="CH211-162119"
/clone_lib="CHORI-211"
1. 3108
/feature="assembly_fragment:01033"
fragment_chain:1
clone_end:SP6
vector_side:left
3209. 32890
/feature="assembly_fragment:01691"
fragment_chain:1
32991. 49822
/feature="assembly_fragment:00994"
fragment_chain:1
49923. 55063
/feature="assembly_fragment:00405"
fragment_chain:1
55164. 111844
/feature="assembly_fragment:01949"
fragment_chain:1
111945. 131419
/feature="assembly_fragment:00155"
fragment_chain:1
131520. 149030
/feature="assembly_fragment:00672"
149131. 159948
/feature="assembly_fragment:00468"
clone_end:T7
vector_side:right"

BASE COUNT 51499 a 27454 c 27877 g 52417 t 701 others
ORIGIN

Query Match 90.5%; Score 19; DB 2; Length 159948;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTGATTGATGACCAG 19
|||||
Db 92696 AAGCTGATTGATGACCAG 92678

RESULT 5
AK098443 1494 bp mRNA linear PRI 15-JUL-2002
LOCUS
DEFINITION Homo sapiens cDNA FLJ25577 fis, clone JTH07710.
ACCESSION AK098443
VERSION AK098443 1 GI:21758456
KEYWORDS oligo cappings; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
Kodaira,H., Furiya,T., Takahashi,M., Kikkawa,B., Omura,Y., Abe,K.,
Kaminara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
Suzuki,Y., Hata,H., Nakagawa,K., Mizuno,S., Morinaga,M.,
Kawamura,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
Nishikawa,T., Sugiyama,A., Kawakami,B., Negai,K., Isogai,T. and
Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1494)
Sugano,S. and Suzuki,Y.
Direct Submission
Submitted (08-JUL-2002) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library

FEATURES
source
Location/Qualifiers
1. 1494
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH07710"
/cell_line="JCR"
/tissue_type="thyroid"
/clone_lib="JTH"
/notes="cloning vector: pME18SFL3"

BASE COUNT 368 a 394 c 292 g 440 t
ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 1494;
Best Local Similarity 90.5%; Pred. No. 5.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGCTGATTGATGACCAGG 21
|||||
Db 1244 AAACCTGGTTGATGACCAGG 1264

RESULT 6
BC048268 1681 bp mRNA linear PRI 31-MAR-2003
LOCUS
DEFINITION Homo sapiens, clone IMAGE:4512974, mRNA.
ACCESSION BC048268
VERSION BC048268.1 GI:29387246
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1681)
Strausberg,R.
Direct Submission
Submitted (07-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgbs-rc@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Contact: nisc.mgc@nih.gov
Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantropop,S., Thomas,P.J., Touchman,J.W.,
Teagueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILLNL at: http://image.llnl.gov
Series: IRAC Plate: 93 Row: d Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
Location/Qualifiers
1. 1681
/organism="Homo sapiens"

FEATURES
source
Location/Qualifiers
1. 1494
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH07710"
/cell_line="JCR"
/tissue_type="thyroid"
/clone_lib="JTH"
/notes="cloning vector: pME18SFL3"

BASE COUNT 368 a 394 c 292 g 440 t
ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 1494;
Best Local Similarity 90.5%; Pred. No. 5.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGCTGATTGATGACCAGG 21
|||||
Db 1244 AAACCTGGTTGATGACCAGG 1264

RESULT 6
BC048268 1681 bp mRNA linear PRI 31-MAR-2003
LOCUS
DEFINITION Homo sapiens, clone IMAGE:4512974, mRNA.
ACCESSION BC048268
VERSION BC048268.1 GI:29387246
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1681)
Strausberg,R.
Direct Submission
Submitted (07-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgbs-rc@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Contact: nisc.mgc@nih.gov
Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantropop,S., Thomas,P.J., Touchman,J.W.,
Teagueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILLNL at: http://image.llnl.gov
Series: IRAC Plate: 93 Row: d Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
Location/Qualifiers
1. 1681
/organism="Homo sapiens"

FEATURES
source
Location/Qualifiers
1. 1494
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH07710"
/cell_line="JCR"
/tissue_type="thyroid"
/clone_lib="JTH"
/notes="cloning vector: pME18SFL3"

BASE COUNT 368 a 394 c 292 g 440 t
ORIGIN

```


Sw:P33479 Sw:Q03391"
/pseudo
/evidence=not_experimental
20312. .20917
/note="CpG island"
/evidence=not_experimental
21213. .21450
/note="L1MC4 repeat: matches 7288. .7525 of consensus"
21645. .21701
/note="L1 repeat: matches 3880. .3935 of consensus"
21704. .21981
/note="L1MC2 repeat: matches 260. .568 of consensus"
21963. .22196
/note="L1MC2 repeat: matches 73. .323 of consensus"
22197. .23067
/note="L1MC2 repeat: matches 1. .897 of consensus"
23068. .23143
/note="L1MC2 repeat: matches 1. .73 of consensus"
23154. .23260
/note="L1 repeat: matches 3759. .3870 of consensus"
23261. .23532
/note="L1MC2 repeat: matches 5757. .6048 of consensus"
23510. .23715
/note="L1MC2 repeat: matches 5411. .5596 of consensus"
23816. .24409
/note="L1MC2 repeat: matches 4796. .5397 of consensus"
24410. .24486
/note="L1MC2 repeat: matches 268. .345 of consensus"
24496. .24630
/note="L1 repeat: matches 4189. .4323 of consensus"
24643. .24784
/note="L1P repeat: matches 3712. .3855 of consensus"
24774. .25930
/note="L1P repeat: matches 4333. .5439 of consensus"
25828. .25903
/note="L1P repeat: matches 5642. .5717 of consensus"
25904. .26345
/note="L1P repeat: matches 5705. .6146 of consensus"
26366. .27087
/note="L1P repeat: matches 5423. .6155 of consensus"
27113. .27322
/note="L1P repeat: matches 1. .212 of consensus"
27323. .28033
/note="L1P repeat: matches 4063. .4797 of consensus"
28034. .28237
/note="L1P repeat: matches 5944. .6158 of consensus"
28238. .28361
/note="L1P repeat: matches 3943. .4063 of consensus"
28361. .28536
/note="L1P repeat: matches 5613. .5792 of consensus"
28537. .28689
/note="L1P repeat: matches 1. .174 of consensus"
28690. .29650
/note="L1P repeat: matches 1. .969 of consensus"
29651. .29765
/note="L1P repeat: matches 174. .282 of consensus"
29848. .30076
/note="L1P repeat: matches 6073. .6226 of consensus"
30077. .30137
/note="L1P repeat: matches 6169. .6226 of consensus"
30383. .31300
/note="CpG island"
/evidence=not_experimental
30658. .31891
/gene="dJ809E13.2"
30658. .31891
/gene="dJ809E13.2"
/product="dJ809E13.2 (putative prenylated protein)"
/note="match: cDNAs: Em:AF052096 Em:AF038168 Em:Y13374
match: ESTs: Em:AI742011 Em:F21403 Em:W76590 Em:AA307887
Em:F21483 Em:N64112 Em:AA421975 Em:AA031541 Em:AA122144
Em:AW410876 Em:AW370386 Em:F37666 Em:AI027000 Em:AA091701

Em:AA302562 Em:AA070647 Em:AA733171 Em:AA076516 Em:C04431
Em:AA864654 Em:AI339190 Em:AI086752 Em:AI346926

Query Match 84.8%; Score 17.8; DB 9; Length 67201;
Best Local Similarity 90.5%; Fred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATGACACGAGG 21
|||
Db 19111 AAACCTGGTGTGACACGAGG 19091
|||

RESULT 8
AC102710
LOCUS
DEFINITION Mus musculus clone RP24-160E14, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
AC102710
VERSION AC102710.3 GI:28395973
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-160E14
Unpublished
2 (bases 1 to 163746)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Grande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamaze,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Ribback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 163746)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,

TITLE
JOURNAL
COMMENT
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (17-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 17, 2003 this sequence version replaced gi:22381702.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L19331
Center clone name: 160 E.14

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 162090 bases at least Q40
Consensus quality: 162587 bases at least Q30
Consensus quality: 162779 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 162946; sum-of-contigs
Quality coverage: 9.4 in Q20 bases; agarose-fp
Quality coverage: 9.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 31737: contig of 31737 bp in length
31738 31837: gap of 100 bp
31838 36585: contig of 4748 bp in length
36586 36586: gap of 100 bp
36586 40932: contig of 4247 bp in length
40933 41032: gap of 100 bp
41033 44827: contig of 3795 bp in length
44828 44927: gap of 100 bp
44928 60753: contig of 15832 bp in length
60750 60760: gap of 100 bp
60760 73424: contig of 12565 bp in length
73425 73425: gap of 100 bp
73425 116324: contig of 42800 bp in length
116325 116424: gap of 100 bp
116425 142656: contig of 26232 bp in length
142657 142757: gap of 100 bp
142757 163746: contig of 20990 bp in length.

FEATURES

source
1. .163746
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-160E14"
/clone_lib="RPC1-24 Male Mouse BAC"
misc_feature
1. .31737
/note="assembly_fragment"
clone_end:SP6
vector_side:left
31838. .36585
/note="assembly_fragment"
36586. .40932
/note="assembly_fragment"
41033. .44827
/note="assembly_fragment"
44928. .60759
/note="assembly_fragment"

misc_feature 60860. .73424
/note="assembly_fragment"
misc_feature 73525. .116324
/note="assembly_fragment"
misc_feature 116425. .142656
/note="assembly_fragment"
misc_feature 142757. .163746
/note="assembly_fragment"
clone_end:T7
vector_side:right

BASE COUNT 44170 a 38132 c 38033 g 43611 t 800 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 163746;
Best Local Similarity 90.5%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCCTGATTGATGACCAGG 21

Db 159477 AGCTTGAATGATGACCAGG 159497

RESULT 9

AC109462/c
LOCUS 174191 bp DNA linear PRI 06-MAR-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-26L20, complete sequence.
ACCESSION AC109462 AC016225
VERSION AC109462.2 GI:19172840

KEYWORDS

HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 174191)

DOE Joint Genome Institute.

AUTHORS

Sequencing of Human Chromosome 16

TITLE

Unpublished

REFERENCE

2 (bases 1 to 174191)

DOE Joint Genome Institute.

AUTHORS

Direct Submission

TITLE

Submitted (04-FEB-2002) Production Sequencing Facility, DOE Joint

JOURNAL

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

3 (bases 1 to 174191)

DOE Joint Genome Institute.

AUTHORS

Direct Submission

TITLE

Submitted (06-MAR-2002) Production Sequencing Facility, DOE Joint

JOURNAL

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On or before Mar 6, 2002 this sequence version replaced
gi:10198460, gi:18483458.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

GenBank flat file format but are available as part

of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

GenBank flat file format but are available as part

of this entry's ASN.1 file.

FEATURES

source

Location/Qualifiers
1. .174191
/organism="Homo sapiens"
/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="16"

/clone="RPL1-26L20"

BASE COUNT 51311 a 36924 c 37255 g 48701 t

ORIGIN

Query Match

Best Local Similarity 84.8%; Score 17.8; DB 9; Length 174191;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTTCATCACCAGG 21

Db 39604 AAGCTGATTTCATCACCAGG 39584

RESULT 10

AC113908/c

LOCUS AC113908 177733 bp DNA linear HTG 15-NOV-2002

DEFINITION Rattus norvegicus clone CH230-396P4, WORKING DRAFT SEQUENCE.

AC113908

VERSION AC113908.9 GI:25008249

KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 177733)

REFERENCE
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Mlosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokaleme,O., Okonou,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,D., Pfamkoch,C.,
Plopper,F., Poinexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 177733)

Worley,K.C.

Direct Submission

Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 177733)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23815563.

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GTDH

Center clone name: CH230-396P4

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 165823 bases at least Q40

Consensus quality: 167723 bases at least Q30

Consensus quality: 168770 bases at least Q20

Estimated insert size: 172340; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 177733: contig of 177733 bp in length.

Location/Qualifiers

1. 177733

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-396P4"

342..1213

/note="clone_boundary"

clone_end:T7

site_

end sequence:B2175527"

complement(176718..177330)

/note="clone_boundary"

clone_end:Sp6

FEATURES

source

misc_feature

misc_feature

site:
 end sequence: BZ175528"
 BASE COUNT 48653 a 37788 c 36881 g 46258 t 8153 others
 ORIGIN
 Query Match 84.8%; Score 17.8; DB 2; Length 177733;
 Best Local Similarity 90.5%; Pred. No. 4.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAGCGTATTGATGACGAGG 21
 Db 72763 AAGCGTATTGATGACGAG 72743
 RESULT 11
 AC128288/c
 LOCUS AC128288 180353 bp DNA linear HTG 20-NOV-2002
 DEFINITION Rattus norvegicus clone CH230-390F9, WORKING DRAFT SEQUENCE.
 ACCESSION AC128288
 VERSION AC128288.3 GI:25137732
 KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 180353)
 Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gibrat, P., Gierke, K., Hill, R., Grady, M., Guerra, W., Guevara, W.,
 Guneratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensu, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwaokeleneh, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
 Popper, F., Poudexter, A., Popovic, D., Primus, E., Pu, L.L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Taylor, T., Thomas, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, R., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 180353)
 Worley, K.C.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 180353)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 COMMENT
 On Nov 20, 2002 this sequence version replaced gi:22856104.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GYDD
 Center clone name: CH230-390F9
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 168606 bases at least Q40
 Consensus quality: 170052 bases at least Q30
 Consensus quality: 171200 bases at least Q20
 Estimated insert size: 171038; sum-of-contigs estimation
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 180353: contig of 180353 bp in length.
 * Location/Qualifiers
 1..180353
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-390F9"
 misc_feature 1..1070
 /note="wgs contig"
 misc_feature 2565..4033
 /note="wgs contig"
 misc_feature 4915..6314
 /note="wgs contig"
 BASE COUNT 48322 a 37960 c 37295 g 48481 t 8295 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 180353;
 Best Local Similarity 90.5%; Pred. No. 4.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATGCACGGG 21
 Db 30172 AAGCTGATTGATGCAGAG 30152

RESULT 12

AC117856/c
 LOCUS AC117856 186118 bp DNA linear HTG 11-OCT-2002
 DEFINITION Rattus norvegicus clone CH230-327H18, WORKING DRAFT SEQUENCE, 3
 unordered pieces.
 ACCESSION AC117856
 VERSION AC117856.4 GI:23814006
 KEYWORDS HTG; HTGS PHASE3; HTGS DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 186118)
 Muzny,D.Marie., Metzker,M.Lee., Abranzone,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gregorogis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorensuwa,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
 Nwakoelameh,O., Okwono,G., Olarpunagoon,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajls,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tincey,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villanana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao S., Dunn,D., von
 Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstock,G. and Gibbs,R.A.
 Direct Submission

TITLE

JOURNAL

REFERENCE 2 (bases 1 to 186118)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL

Submitted (11-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 186118)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Oct 11, 2002 this sequence version is replaced gi:21745720.

The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GTTF

Center clone name: CH230-327H18

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 169552 bases at least Q40

Consensus quality: 171519 bases at least Q30

Consensus quality: 172461 bases at least Q20

Estimated insert size: 173910; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOT: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOT: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 182244: contig of 182244 bp in length

* 182345 182344: gap of unknown length

* 182345 182442: contig of 1898 bp in length

* 184243 184342: gap of unknown length

* 184343 186118: contig of 1776 bp in length.

FEATURES

source

1. 186118

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clones="CH230-327H18"

1. 1031

/note="wgs end extension"

clone_end:T7"

complement(5228..6123)

/note="clone boundary"

clone_end:T7"

site:MboI

end_sequence:RXAOI45TJ"

171082..171985

/note="clone_boundary"

misc_feature

misc_feature

misc_feature

```

clone end:Sp6
site:Mbol
end sequence:RXAOL45TV"
misc_feature
179115..180519
/Note="wgs_end_extension
clone end:Sp6"
181021..182244
/Note="wgs_end_extension
clone end:Sp6"
BASE COUNT 49746 a 33747 c 35829 g 53877 t 12919 others
ORIGIN
Query Match 84.8%; Score 17.8; DB 2; Length 186118;
Best Local Similarity 90.5%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCCCTGATGATGACACGGG 21
|||||
Db 164995 AGCCCTGATGATGACACGGT 164975

RESULT 13
AC111040
LOCUS
DEFINITION Mus musculus clone RP24-536C5, WORKING DRAFT SEQUENCE, 11 ordered
pieces.
ACCESSION AC111040
VERSION AC111040.4 GI:22381727
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (Bases 1 to 194454)
Birren, B., Nusbaum, C. and Lander, E.
MUS musculus, clone RP24-536C5
Unpublished
TITLE Mus musculus, clone RP24-536C5
JOURNAL
REFERENCE
2 (bases 1 to 194454)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, K., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kellis, C., LaRocque, K., Lamazates, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McSwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194454)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepe, K., Collamore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:20806337.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19423
Center clone name: 536_C_5
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191374 bases at least Q40
Consensus quality: 192313 bases at least Q30
Consensus quality: 192832 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 193454; sum-of-contigs
Quality coverage: 10.6 in Q20 bases; agarose-fp
Quality coverage: 10.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 3686: contig of 3686 bp in length
* 3687 3787: gap of 100 bp
* 3787 11203: contig of 7417 bp in length
* 11204 11303: gap of 100 bp
* 11304 22195: contig of 10891 bp in length
* 22195 32225: contig of 9931 bp in length
* 32225 32325: gap of 100 bp
* 32325 32326: contig of 100 bp
* 32326 39995: contig of 7670 bp in length
* 39995 40095: gap of 100 bp
* 40095 57751: contig of 17656 bp in length
* 57751 57851: gap of 100 bp
* 57851 69354: contig of 11503 bp in length
* 69354 69455: gap of 100 bp
* 69455 87475: contig of 18021 bp in length
* 87475 87575: gap of 100 bp
* 87575 118208: contig of 30633 bp in length
* 118208 118308: gap of 100 bp
* 118308 158787: contig of 40479 bp in length
* 158787 158888: gap of 100 bp
* 158888 194454: contig of 35567 bp in length.
* Location/Qualifiers
* 1. 194454
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP24-536C5"
* /clone_lib="RPCI-24 Male Mouse BAC"

```

misc_feature 1. 3686
/note="assembly_fragment"
misc_feature 3787. 11203
/note="assembly_fragment"
misc_feature 11304. 22194
/note="assembly_fragment"
misc_feature 22295. 32225
/note="assembly_fragment"
misc_feature 32326. 33995
/note="assembly_fragment"
misc_feature 40096. 57751
/note="assembly_fragment"
misc_feature 57852. 69354
/note="assembly_fragment"
misc_feature 69455. 87475
/note="assembly_fragment"
misc_feature 87576. 118208
/note="assembly_fragment"
misc_feature 118309. 158787
/note="assembly_fragment"
misc_feature 158888. 194454
/note="assembly_fragment"

BASE COUNT 53011 a 45075 c 45339 g 50027 t 1002 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 194454;
Best Local Similarity 90.5%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATGACGAGG 21

Db 134728 AAGCTGATTGATGACGAGG 134748

RESULT 14

AL691424/c
LOCUS AL691424 194520 bp DNA linear ROD 26-JUL-2002
DEFINITION Mouse DNA sequence from clone RP23-219X12 on chromosome X, complete sequence.
ACCESSION AL691424
VERSION AL691424.7 GI:22002699
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194520)
Wray.P.
Direct Submission
Submitted (26-JUL-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Jul 29, 2002 this sequence version replaced gi:21912679.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-219K12 is from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6.

FEATURES
source
1. 194520
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-219K12"
/clone_lib="RPCI-23"
BASE COUNT 55728 a 38763 c 37184 g 62845 t
ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 194520;
Best Local Similarity 90.5%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGCTGATTGATGACGAGG 21

Db 42410 ATGCTGATTGATGACGAGG 42390

RESULT 15

AC123058
LOCUS AC123058 194636 bp DNA linear ROD 02-OCT-2002
DEFINITION Mus musculus chromosome 19 clone RP23-188J8, complete sequence.
ACCESSION AC123058
VERSION AC123058.3 GI:23462991
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194636)
McPherson, J.D. and Waterston, R.H.
The sequence of Mus musculus clone
Unpublished
REFERENCE 2 (bases 1 to 194636)
McPherson, J.D. and Waterston, R.H.
Direct Submission
TITLE Direct Submission
AUTHORS McPherson, J.D. and Waterston, R.H.
JOURNAL Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 194636)
McPherson, J.D. and Waterston, R.H.
Direct Submission
TITLE Direct Submission
AUTHORS McPherson, J.D. and Waterston, R.H.
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 194636)
McPherson, J.D. and Waterston, R.H.
Direct Submission
TITLE Direct Submission
AUTHORS McPherson, J.D. and Waterston, R.H.
JOURNAL Submitted (02-OCT-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Oct 2, 2002 this sequence version replaced gi:22475873.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0188J08

FEATURES
source
1. 194636
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

```

/chromosome="19"
/clone="RP23-188J8"
BASE COUNT 50027 a 45155 c 46196 g 53258 t
ORIGIN
Query Match 84.8%; Score 17.8; DB 10; Length 194636;
Best Local Similarity 90.5%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGCCTGATTGATGACGAGG 21
Db 50616 AAGCTGAATGATGACGAGG 50636

```

Search completed: February 19, 2004, 22:22:05
Job time : 1601.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2004, 01:49:48 ; Search time 197 Seconds
(without alignments)
392.672 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCTCACCCTTGCTCA 21

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 490076

Minimum DB seq length: 0
Maximum DB seq length: 21

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	14.8	70.5	20	13	US-10-171-319-48
2	14.8	70.5	20	13	US-10-171-319-73
3	14.2	67.6	20	13	US-09-865-879-29
c 4	14	66.7	20	9	US-09-735-995-47
5	13.4	63.8	21	12	US-10-349-143-10433
6	13.2	62.9	20	11	US-09-953-047-23
c 7	13.2	62.9	20	15	US-10-128-714-4014
c 8	12.8	61.0	17	13	US-10-061-201-866
c 9	12.8	61.0	17	13	US-10-061-201-867
c 10	12.8	61.0	18	11	US-09-992-331-7
c 11	12.8	61.0	18	16	US-10-262-313-7
c 12	12.8	61.0	20	10	US-09-791-406-63
c 13	12.8	61.0	21	10	US-09-922-364A-35
c 14	12.8	61.0	21	10	US-09-254-590-35
c 15	12.8	61.0	21	12	US-10-349-143-9636

c 16	12.8	61.0	21	14	US-10-115-695-35
c 17	12.8	61.0	21	15	US-10-116-561-35
c 18	12.8	61.0	21	15	US-10-115-671-35
c 19	12.8	61.0	21	15	US-10-115-415-35
c 20	12.8	61.0	21	15	US-10-116-260-35
c 21	12.8	61.0	21	15	US-10-115-688-35
22	12.6	60.0	20	11	US-09-912-724-42
23	12.6	60.0	20	13	US-10-321-555-2
c 24	12.6	60.0	20	13	US-10-259-451-12
c 25	12.6	60.0	21	12	US-10-418-182-126
c 26	12.4	59.0	17	13	US-10-061-201-864
c 27	12.4	59.0	17	13	US-10-061-201-865
c 28	12.4	59.0	19	10	US-09-966-147-22
c 29	12.4	59.0	19	13	US-10-374-469-22
c 30	12.2	58.1	17	9	US-09-866-108-8077
c 31	12.2	58.1	17	9	US-09-866-108-8078
c 32	12.2	58.1	17	9	US-09-866-108-8398
c 33	12.2	58.1	17	9	US-09-866-108-8399
34	12.2	58.1	17	11	US-09-740-332-2130
35	12.2	58.1	17	13	US-09-817-879-2130
36	12.2	58.1	18	12	US-10-349-143-11706
c 37	12.2	58.1	20	11	US-09-917-963-85
38	12	57.1	18	10	US-09-870-956-11
39	12	57.1	19	13	US-10-224-005-38
c 40	12	57.1	19	13	US-10-224-005-199
c 41	12	57.1	20	11	US-09-888-326-807
c 42	12	57.1	20	11	US-09-776-479-446
c 43	12	57.1	20	13	US-10-002-623-606
c 44	12	57.1	20	15	US-10-112-653-426
c 45	12	57.1	20	15	US-10-017-995-446

ALIGNMENTS

RESULT 1

US-10-171-319-48/c
; Sequence 48, Application US/10171319
; Publication No. US20030157633A1
; GENERAL INFORMATION:
; APPLICANT: Ardem Patapoutian
; APPLICANT: Andrea Peier
; APPLICANT: Peter McIntyre
; APPLICANT: Stuart Bevan
; APPLICANT: Chuansheng Song
; APPLICANT: Pamposh Ganju
; TITLE OF INVENTION: VANILLOID RECEPTOR-RELATED NUCLEIC ACIDS
; TITLE OF INVENTION: AND POLYPEPTIDES
; FILE REFERENCE: 4-32048A
; CURRENT APPLICATION NUMBER: US/10/171,319
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 60/297,835
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/351,238
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/352,914
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/357,161
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/381,086
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/381,739
; PRIOR FILING DATE: 2002-05-16
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide primer
US-10-171-319-48

Sequence 35, Appl
Sequence 35, Appl
Sequence 35, Appl
Sequence 35, Appl
Sequence 35, Appl
Sequence 35, Appl
Sequence 35, Appl
Sequence 42, Appl
Sequence 2, Appl
Sequence 12, Appl
Sequence 126, Appl
Sequence 864, Appl
Sequence 865, Appl
Sequence 22, Appl
Sequence 22, Appl
Sequence 8077, Ap
Sequence 8078, Ap
Sequence 8398, Ap
Sequence 8399, Ap
Sequence 2130, Ap
Sequence 11706, A
Sequence 85, Appl
Sequence 11, Appl
Sequence 38, Appl
Sequence 199, Appl
Sequence 807, Appl
Sequence 446, Appl
Sequence 606, Appl
Sequence 426, Appl
Sequence 446, Appl

```

; SEQ ID NO 29
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Anisense primer for p28 alpha
US-09-865-879-29

Query Match          67.6%; Score 14.2; DB 13; Length 20;
Best Local Similarity 84.2%; Pred. No. 7.3e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TCATCTCTCACCCCTGTGCCT 19
        ||||| ||||| |||||
Db       2 TCATCTCTCCCCCTTCTCT 20

RESULT 4
US-09-735-995-47/c
; Sequence 47, Application US/097355995
; Patent No. US20010034024A1
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; TITLE OF INVENTION: SYNDROME GENE
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/735,995
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 09/226,012
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-995-47

Query Match          66.7%; Score 14; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.9e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 CCTCACCCCTTGTC 18
        ||||| ||||| |||||
Db       16 CCTCACCCCTTGTC 3

RESULT 5
US-10-349-143-10433
; Sequence 10433, Application US/10349143
; Publication No. US20040005584A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
; FILE REFERENCE: GENSET.020CP1
; CURRENT APPLICATION NUMBER: US/10/349,143
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US/09/422,978
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/298,850
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/109,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/082,614
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 10433
; LENGTH: 21
; TYPE: DNA

```

```
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1..21
; OTHER INFORMATION: downstream amplification primer 99-11851 for SEQ 2568, in complement
US-10-349-143-10433

Query Match      63.8%; Score 13.4; DB 12; Length 21;
Best Local Similarity 93.3%; Pred. No. 1.6e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 CTCACCCCTTGTCCTC 20
Db 7 CTCACCCCTTGTCCTC 21

RESULT 6
US-09-953-047-23
; Sequence 23, Application US/09953047
; Publication No. US20030087854A1
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF FIBROBLAST GROWTH FACTOR RECEPTOR 3 EXPRESSION
; FILE REFERENCE: RTS-0157
; CURRENT APPLICATION NUMBER: US/09/953,047
; CURRENT FILING DATE: 2001-09-10
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 23
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-953-047-23

Query Match      62.9%; Score 13.2; DB 11; Length 20;
Best Local Similarity 83.3%; Pred. No. 2e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 TCCTCACCCCTTGTCCTCA 21
Db 2 TCCACACCTGTGTCTCTCA 19

RESULT 7
US-10-128-714-4014/c
; Sequence 4014, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: Methods of Use
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 4014
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-4014

Query Match      62.9%; Score 13.2; DB 15; Length 20;
Best Local Similarity 83.3%; Pred. No. 2e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CATCCTCACCCCTTGTCCT 19
Db 19 CATACCCATCCTTGTCCT 2

RESULT 8
US-10-061-201-866/c
; Sequence 866, Application US/10061201
; Publication No. US20030166229A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: HUMAN POSH-LIKE PROTEIN 1
; FILE REFERENCE: PB0178
; CURRENT APPLICATION NUMBER: US/10/061,201
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/328,205
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 4162
; SOFTWARE: Asomica Sequence Listing Engine
; SEQ ID NO 866
; LENGTH: 17
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-061-201-866

Query Match      61.0%; Score 12.8; DB 13; Length 17;
Best Local Similarity 87.5%; Pred. No. 3e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 TCCTCACCCCTTGTCCT 19
Db 17 TTCACACTCTTGTCCT 2

RESULT 9
US-10-061-201-867/c
; Sequence 867, Application US/10061201
; Publication No. US20030166229A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: HUMAN POSH-LIKE PROTEIN 1
; FILE REFERENCE: PB0178
; CURRENT APPLICATION NUMBER: US/10/061,201
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666
```


;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 09/864,761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/328,205
;; PRIOR FILING DATE: 2001-10-10
;; NUMBER OF SEQ ID NOS: 4162
;; SOFTWARE: Acomica Sequence Listing Engine
;; SEQ ID NO 867
;; LENGTH: 17
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-061-201-867

Query Match 61.0%; Score 12.8; DB 13; Length 17;
Best Local Similarity 87.5%; Pred. No. 3e+04; 2; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 2;

QY 4 TCCTCACCCCTTGCT 19
DB 16 TTCTCACTCTTGCT 1

RESULT 10
US-09-992-331-7/c
;; Sequence 7, Application US/0992331
;; Publication No. US20030022186A1
;; GENERAL INFORMATION:
;; APPLICANT: FEDER, JOHN N.
;; APPLICANT: MINTIER, GABE
;; APPLICANT: RAMANATHAN, CHANDRA S.
;; APPLICANT: HAWKEN, DONALD R.
;; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBMY18,
;; TITLE OF INVENTION: EXPRESSED HIGHLY IN PITUITARY GLAND AND COLON CARCINOMA
;; TITLE OF INVENTION: CELLS
;; FILE REFERENCE: D0048NP
;; CURRENT APPLICATION NUMBER: US/09/992,331
;; CURRENT FILING DATE: 2001-11-14
;; PRIOR APPLICATION NUMBER: 60/308,540
;; PRIOR FILING DATE: 2001-07-27
;; PRIOR APPLICATION NUMBER: 60/261,782
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: 60/248,483
;; PRIOR FILING DATE: 2000-11-14
;; NUMBER OF SEQ ID NOS: 45
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 7
;; LENGTH: 18
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: HGPBMY8
;; OTHER INFORMATION: anti-sense primer
US-09-992-331-7

Query Match 61.0%; Score 12.8; DB 11; Length 18;
Best Local Similarity 87.5%; Pred. No. 3e+04; 2; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 2;

QY 2 CATCTCACCCCTTGTC 17

DB 16 CATCTCACCCCTTGTC 1
RESULT 11
US-10-262-313-7/c
;; Sequence 7, Application US/10262313
;; Publication No. US20030129653A1
;; GENERAL INFORMATION:
;; APPLICANT: Bristol-Myers Squibb Company
;; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBMY18, EXPRESSED HIGH
;; TITLE OF INVENTION: PITUITARY GLAND AND COLON CARCINOMA CELLS
;; FILE REFERENCE: D0048 CIP
;; CURRENT APPLICATION NUMBER: US/10/262,313
;; CURRENT FILING DATE: 2002-09-30
;; PRIOR APPLICATION NUMBER: U.S. 09/992,331
;; PRIOR FILING DATE: 2001-11-14
;; NUMBER OF SEQ ID NOS: 50
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 7
;; LENGTH: 18
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-262-313-7

Query Match 61.0%; Score 12.8; DB 16; Length 18;
Best Local Similarity 87.5%; Pred. No. 3e+04; 2; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 2;

QY 2 CATCTCACCCCTTGTC 17
DB 16 CATCTCACCCCTTGTC 1

RESULT 12
US-09-791-406-63
;; Sequence 63, Application US/09791406
;; Patent No. US20020147165A1
;; GENERAL INFORMATION:
;; APPLICANT: C. Frank Bennett
;; APPLICANT: Robert Rothlein
;; APPLICANT: Takashi Kei Kishimoto
;; APPLICANT: Lex M. Cowsett
;; TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
;; FILE REFERENCE: RTS-0097
;; CURRENT APPLICATION NUMBER: US/09/791,406
;; CURRENT FILING DATE: 2001-02-22
;; NUMBER OF SEQ ID NOS: 89
;; SEQ ID NO 63
;; LENGTH: 20
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Antisense Oligonucleotide
US-09-791-406-63

Query Match 61.0%; Score 12.8; DB 10; Length 20;
Best Local Similarity 87.5%; Pred. No. 2.9e+04; 2; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 2;

QY 6 CTCACCCCTTGCTCTCA 21
DB 1 CTCATCTTTGCTCTCA 16

RESULT 13
US-09-922-364A-35/c
;; Sequence 35, Application US/09922364A
;; Patent No. US2002015531A1
;; GENERAL INFORMATION:
;; APPLICANT: Adelman, John P.
;; APPLICANT: Maylie, James
;; APPLICANT: Bond, Chris T.

; Silvia, Christopher P.
; TITLE OF INVENTION: Small and Intermediate Conductance,
; Calcium-Activated Potassium Channels and Uses
; Thereof
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/922.364A
; FILING DATE: 03-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/254,590
; FILING DATE: 10-Mar-1999
; APPLICATION NUMBER: US 60/026,451
; FILING DATE: 11-SEP-1996
; APPLICATION NUMBER: US 60/040,052
; FILING DATE: 07-MAR-1997
; APPLICATION NUMBER: US 60/045,233
; FILING DATE: 17-APR-1997
; APPLICATION NUMBER: WO PCT/US97/16033
; FILING DATE: 10-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 014210-000730US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-922-364A-35

Query Match 61.0%; Score 12.8; DB 10; Length 21;
Best Local Similarity 87.5%; Pred. No. 2.9e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CCTCACCCTTGCTCTC 20
Db 19 CCTCAGCTTGGCTC 4

RESULT 14
US-09-254-590-35/c
; Sequence 35, Application US/09254590
; Patent No. US20020165379A1
; GENERAL INFORMATION:
; APPLICANT: Adelman, John P.
; Maylie, James
; Bond, Chris I.
; Silvia, Christopher P.
; TITLE OF INVENTION: Small and Intermediate Conductance,
; Calcium-Activated Potassium Channels and Uses Thereof
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/254,590
; FILING DATE: 10-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,451
; FILING DATE: 11-SEP-1996
; APPLICATION NUMBER: US 60/040,052
; FILING DATE: 07-MAR-1997
; APPLICATION NUMBER: US 60/045,233
; FILING DATE: 17-APR-1997
; APPLICATION NUMBER: WO PCT/US97/16033
; FILING DATE: 10-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 014210-000730US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-254-590-35

Query Match 61.0%; Score 12.8; DB 10; Length 21;
Best Local Similarity 87.5%; Pred. No. 2.9e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CCTCACCCTTGCTCTC 20
Db 19 CCTCAGCTTGGCTC 4

RESULT 15
US-10-349-143-9636/c
; Sequence 9636, Application US/10349143
; Publication No. US20040005584A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
; FILE REFERENCE: GENSET.020CP1
; CURRENT APPLICATION NUMBER: US/10/349,143
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US/09/422,978
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/298,850
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/109,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/082,614
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 9636
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Homo Sapiens

```
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1..21
; OTHER INFORMATION: downstream amplification primer 99-6261 for SEQ 1771, in compleme
US-10-349-143-9636
Query Match      61.0%; Score 12.8; DB 12; Length 21;
Best Local Similarity 87.5%; Pred. No. 2.9e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      5 CCTCACCCCTTGTCTC 20
        |||||
Db      21 CCTCACCCCTTCTC 6
```

Search completed: February 20, 2004, 04:09:30
Job time : 198 secs